Free Cooling-Aware Dynamic Power Management for Green Datacenters

Jungsoo Kim
Embedded Systems Lab (ESL), EPFL
Lausanne, Switzerland
jungsoo.kim@epfl.ch

Martino Ruggiero
Embedded Systems Lab (ESL), EPFL
Lausanne, Switzerland
martino.ruggiero@epfl.ch

David Atienza
Embedded Systems Lab (ESL), EPFL
Lausanne, Switzerland
david.atienza@epfl.ch

***Abstract*—Free cooling, i.e., directly using outside cold air and/or water to cool down datacenters, can provide significant power savings of datacenters. However, due to the limited cooling capability, which is tightly coupled with climate conditions, free cooling is currently used only in limited locations (e.g., North Europe) and periods of the year. Moreover, the applicability of free cooling is further restricted along with the conservative assumption on workload characteristics and the virtual machine (VM) consolidation technique as they require to provision higher cooling capability. This paper presents a dynamic power management scheme, which extends the applicability of free cooling by judiciously consolidating VMs exploiting time-varying workload characteristics of datacenter as well as climate conditions, in order to minimize the power consumption of the entire datacenter while satisfying service-level agreement (SLA) requirements. Additionally, we propose the use of a *receding horizon control* scheme in order to prevent frequent cooling mode transitions. Experimental results show that the proposed solution provides up to 25.7% power savings compared to conventional free cooling decision schemes, which uses free cooling only when the outside temperature is lower than predefined threshold temperature.**

***Keywords*—Architectures for green computing, Power aware architectures, Virtualization and virtual machines, Energy efficient HPC systems**

I. Introduction

Within the total power use of datacenters, a significant fraction is devoted to the cooling infrastructure to maintain servers within their safe operating conditions, e.g., $64.4°F \sim 80.6°F$ of temperature and $41.9°F \sim 59°F$ of dew point [1]. According to US EPA report [2], power usage effectiveness (PUE), defined as the ratio of the total power consumed by a datacenter with respect to the power consumed by servers, amounts to 1.9 on average for current datacenters in the world. This means that for every watt of power consumed in the computing equipment, an additional 0.9W of power is spent as power overhead (cooling, power distribution, etc.).

In order to reduce the power consumption of cooling facilities, various solutions have been proposed, e.g., hot-/cold-aisle layout [3], dynamic adjustment of thermostat of server room [4]. Nonetheless, these practices reach a reported PUE in the order of 1.4 [2], which is still far from the energy minimal target (PUE$\simeq$1.0). A recent approach to improve energy efficiency in datacenters is the concept of free cooling, which relies on the use of outside cold air and/or water for cooling instead of electricity (namely, electrical cooling) [5]. Free cooling operates on the principle that during cool/cold weather conditions datacenter cooling loads can be served with chilled water produced by the cooling tower alone, entirely bypassing the energy-costly chillers. Thus, free cooling reduces or eliminates chiller power consumption while efficiently maintaining strict temperature and humidity requirements. This is a promising architectural innovation for datacenter cooling infrastructure that can enable PUE to approach values near 1.0 [6]. However, despite the promising advantages on cooling energy efficiency, the fundamental shortcoming of free cooling is its limited applicability, as it can only be used in a very limited set of geographical locations because the cooling capability is tightly coupled with climate conditions (e.g., temperature and humidity).

Hybrid cooling [6]–[12], which provisions back-up cooling infrastructure along with free cooling, is an intuitive solution to extend the usability of free cooling. It switches between free and electrical cooling according to outside climate condition. For instance, if the outside temperature is lower than a certain threshold, free cooling is used; otherwise, chiller-based electric cooling is employed. However, the time period when free cooling can be used is still very limited as the threshold temperature has to be set typically very low, e.g., $8°C$ as in [12], in order to cope with even the worst-case workload scenario of datacenter. However, the worst-case scenario rarely happens [13]. Thus, such conservative decision of cooling mode without considering workload variation deprives the chance of using free cooling. Furthermore, when it comes along with existing solutions to reduce computing power consumption, especially virtual machine (VM) consolidation [14]–[19], the applicability of free cooling becomes drastically reduced because the solution leads to higher operating temperature of actively running servers, as it increases the resource utilization by consolidating VMs into smaller number of servers. However, based on our observations, VMs are consolidated in such a way that the maximum power consumption of active servers is less than the level where the cooling capability provided by free cooling is sufficient, we

This work described in this paper has been partially supported by the PMSM: CT Monitoring research grant for ESL-EPFL funded by Credit Suisse AG.

978-1-4673-2362-8/12/$31.00 ©2012 IEEE

can save the power consumption by extending the time period of using free cooling.

Motivated by these observations above, in this paper, we propose a dynamic power management solution for datacenters having a hybrid cooling architecture so as to reduce the power consumption while satisfying service-level agreement (SLA) requirements by extending the usability of free cooling. To achieve this goal, we determine the optimal pair of cooling mode and maximum power consumption of active servers (namely, *power capping*) by considering time-varying and uncertain characteristics of both outside temperature and workload characteristics of datacenters. Furthermore, in order to jointly optimize the overhead caused by switching cooling mode, we propose the use of a novel receding horizon control scheme which periodically determine the optimal pair considering the predictive sequence of cooling mode transition. Our experimental results show that the proposed scheme extends the period of using free cooling, thereby, providing up to 25.7% power savings compared to conventional cooling mode decision scheme based on a predefined fixed threshold temperature.

This paper is organized as follows. In Section II, we define our target datacenter architecture. In Section III, we describe the target problem definition and our proposed solution. Next in Section IV, we explore the trade-offs of our joint optimization procedure between the target cooling power minimization and the possible server utilization levels. Then, in Section V, we present our experimental results and, finally, Section VI summarizes the main conclusions of this work.

II. Target System Architecture

A. Target architecture

The target datacenter mainly consists of computing and cooling parts. Fig. 1(a) provides an overview of the we address in this paper. The computing part consists of five main components: 1) an application queue in which user requests queue up for processing; 2) a resource manager which distributes user requests to virtual machines (VMs); 3) N_{vm} homogeneous VMs in each of which application is running along with operating system (OS); iv) a VM scheduler which allocates VMs to servers; and 5) N_{pm} homegenous servers or physical machines (PMs). Note that as many VMs are provisioned in a single server, response time increases drastically due to resource conflict, e.g., pipeline, cache, memory, etc., especially after exceeding certain utilization threshold level, u_{pm}^{max} [20]. Thus, we assume that VM scheduler allocates VMs to servers such that the server utilization does not exceed u_{pm}^{max}, e.g., 0.8 in [20].

Regarding the cooling part, we target a hybrid cooling architecture consisting of Computer Room Air Handler (CRAH), chiller, and cooling tower as shown in Fig. 1(b). CRAH transfers the heat flux out of the server room and provides a new source of cooled air by exchanging heat with chilled

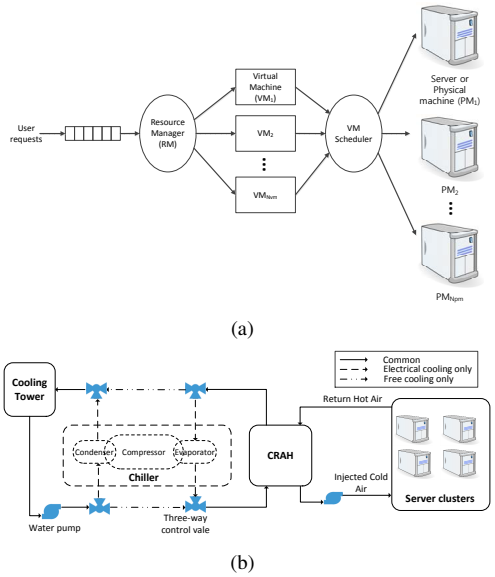

Figure 1. Target datacenter architecture: (a) computing and (b) cooling parts.

water provided by chiller or outside cold water. When the chilled water is provided by the chillers, we call the cooling mode *electrical cooling*; otherwise, *free cooling*. In this figure, a solid line represents a common path activated in both electrical and free cooling; while a dotted and a double dotted line represent paths exclusively activated in electrical and free cooling, respectively. In this work, we adjust the VM assignment and cooling mode at every t_{opt} or when SLA violation exceeds a predefined threshold level.

B. Power and temperature model

1) Computing system: A server has multiple power modes, $m_{pm_i} \in \{active, idle, sleep\}$ [21], each of which has different power consumption and response time. We model the power consumption of each power mode as follows [22]:

$$P_{pm_i} = \begin{cases} P_{pm}^{static} + P_{pm}^{dyn} u_{pm_i} & \text{, if } m_{pm_i} = active \\ P_{pm}^{idle} & \text{, if } m_{pm_i} = idle \\ P_{pm}^{sleep} & \text{, if } m_{pm_i} = sleep \end{cases} \quad (1)$$

where P_{pm}^{static} and P_{pm}^{dyn} are constants that model the static and dynamic power consumption when a server is in active mode. Then, u_{pm_i} represents CPU utilization of the server, and P_{pm}^{idle} and P_{pm}^{sleep} are constants representing the power consumption at idle and sleep mode, respectively. Therefore, the power consumed by server clusters in a datacenter can be calculated as the sum of power consumption of the servers, as follows:

$$P_{cl} = \sum_{i=1}^{N_{pm}} P_{pm_i} \quad (2)$$

Then, we model the (steady-state) temperature of each server based on the well-known duality between thermal and electrical phenomena [23], namely:

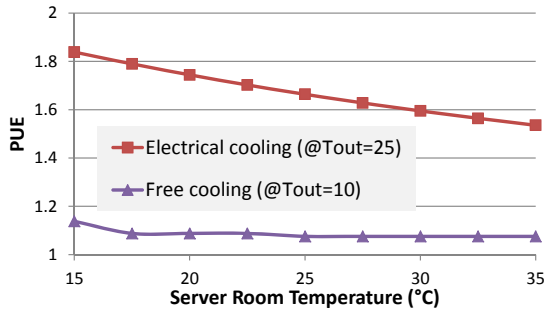

Figure 2. Power Usage Effectiveness (PUE) in electrical and free cooling as power consumption of server varies.

$$T_{pm_i} = T_{room} + R_{pm} \cdot P_{pm_i} \tag{3}$$

where T_{room} is the ambient temperature of the server room. And R_{pm} is the thermal resistance between die and air, which can be calculated as the sum of thermal resistances of the package, thermal interface material (TIM), and heat sink [24].

2) Cooling system: According to the definition of PUE, the power consumed by the cooling system, P_{co}, can be calculated as follows:

$$P_{co} = (PUE - 1) \cdot P_{cl} \tag{4}$$

In fact, PUE is usually modeled with complex equations based on thermo-fluid principles [24] [25]. However, based on our analysis of real datacenter setups of our industrial partners in this work, we have observed that an alternative procedure can be used, where PUE mainly depends on the temperature set-point of server room (T_{room}), outside temperature (T_{out}), and total power consumed by servers (P_{cl}). Moreover, T_{room} is the dominant factor compared to the others. Thus, we can simply characterize PUE with respect to T_{room}. Fig. 2 shows PUE with respect to T_{room} when $R_{pm} = 0.5W/°C$. As shown in this figure, the PUE of electrical and free cooling ranges 1.53∼1.83 and 1.08∼1.14, respectively. Assuming that T_{room} is set to the highest temperature of which servers in active mode can satisfy the maximum temperature limit, i.e., T_{pm}^{max}, we can model PUE as a function of P_{pm} with Eqn. (3). By matching the results shown in Fig. 2 and Eqn. (3), we can approximate the PUE with a relatively simple form, namely:

$$PUE = a_1 P_{pm}^2 + a_2 P_{pm} + a_3 \tag{5}$$

where a_1, a_2, and a_3 are curve fitting parameters. In the case of electrical and free cooling, the sets we have obtained for $\{a_1, a_2, a_3\}$ are $\{3.32 \times 10^{-5}, -9.45e \times 10^{-4}, 1.30\}$ and $\{0, 0, 1.08\}$, respectively. Then, the maximum (average) root mean square (RMS) error amounts to 4.38% (0.76%) and 0.56%(0.56%), respectively.

Finally, the Temperature of the server room, T_{room}, depends on CRAH efficiency, ϵ_{CRAH}, which is defined as follows [25]:

$$\epsilon_{CRAH} = \frac{T_{CRAH}^{air} - T_{room}}{T_{CRAH}^{air} - T_{CRAH}^{water}} \tag{6}$$

In this equation, T_{CRAH}^{air} represents temperatures of air exhausted from server room; T_{CRAH}^{water} is the temperature of chilled water flowing into the CRAH, which corresponds to the set-point of chiller and outside temperature when electrical and free cooling is used, respectively. Since ϵ_{CRAH} is always less than 1, T_{room} is always higher than T_{CARH}^{water}.

C. Datacenter workload model

The workload characteristics of the datacenter depend on the pattern of users' requests demanded to the datacenter computing systems, which can be characterized with two different time scales: 1) microscopic (less than few seconds) and 2) macroscopic (few tens of minutes to hours). At the microscopic scale, the characteristics of user requests depend on burstiness of traffic and arrival patterns. As presented in [26], we model the characteristics of users' request at the microscopic scale with 1) ON/OFF periods and 2) inter-arrival time between two consecutive requests during ON period. ON period is defined as the longest continual period during which all the request inter-arrival times are smaller than predefined value. Accordingly, OFF period is defined as a period between two on periods. As presented in [26], ON/OFF period and inter-arrival time are time-varying and uncertain while each of them forms lognormal distribution. At the macroscopic scale, the characteristics of users' requests are distinctly different over time while the global pattern has a strong correlation with adjacent time periods as well as the same period in different days [27].

Therefore, in our optimization problem, we model the uncertain workload characteristics with statistical parameters: 1) maximum user request, 2) average user request, 3) standard deviation, and 4) correlation with other adjacent time periods.

III. Problem Definition and Optimization Approach

The problem we are trying to tackle is two-fold, namely, determining both the 1) cooling mode and 2) VM placement such that the power consumption of datacenter, i.e., $P_{dc} = P_{cl} + P_{co}$, and the overhead caused by cooling mode transition are jointly minimized while satisfying temperature and SLA requirements. Thus, we can formulate the optimization problem as follows:

$$\text{Find} \quad \chi = \{m_{co}, [b_{i,j}]_{N_{pm} \times N_{vm}}\} \tag{7}$$

$$\text{Minimize} \quad J_{dc} = P_{cl} + P_{co} + O_{tr} \tag{8}$$

$$\text{Subject to} \quad T_{pm_i} \leq T_{pm}^{max}, \text{where } 1 \leq i \leq N_{pm} \tag{9}$$

$$Pr(t_{act} > t_{req}) \leq (1 - \beta) \tag{10}$$

In this formulation, m_{co} represents datacenter cooling mode: '1' when free cooling is selected, otherwise '0'; $b_{i,j}$ is a binary variable representing VM placement: '1' when vm_j is mapped into pm_i; J_{dc} is an objective function consisting of power consumption of datacenter, i.e., $P_{dc} = P_{cl} + P_{co}$, and overhead caused by switching cooling mode, i.e., O_{tr}; T_{pm_i} and T_{pm}^{max} represent temperature of i-th server (or physical machine) and

the maximum temperature constraint of servers, respectively. Then, t_{act} and t_{req} are actual and required execution time, respectively, and $Pr(t_{act} > t_{req})$ represents the probability when t_{act} is larger than t_{req}; β is SLA requirement.

This optimization problem can be translated into a bin-packing problem with variable bin size by exploiting the analogy between a bin and a server because, for a given bin size (analogy with threshold of server utilization), the power consumption is minimized when the number of bins (analogy with the number of active servers in which VMs are assigned) is minimized, i.e., server consolidation. Hence, the bin size, i.e., the threshold of server utilization, depends on m_{co} as well as T_{out}. However, due to the interdependency between m_{co} and $b_{i,j}$'s, the solution complexity is even higher than conventional bin-packing problem. To reduce the solution complexity, we propose a two-phase solution to the previously presented problem. First, we determine a power-optimal pair of $\{m_{co}, u^{th}_{pm}\}$ such that J_{dc} is minimized while satisfying temperature and performance requirements assuming that ideal server consolidation[a] is applied, i.e., utilization of every active server equals to u^{th}_{pm} while others are '0'. Second, we assign VMs to servers such that the number of servers where VMs are allocated is minimized while total utilization of every server does not exceed u^{th}_{pm}. In this work, we focus on the first step while simply applying existing heuristics, e.g., first-fit and best-fit, etc. [18], in the second step. Moreover, in order to achieve further improvement by considering time-varying characteristics of T_{out} and the user requests, we iterate the optimization procedure at every predefined time interval, t_{opt}.

IV. Multi-objective trade-offs exploration between cooling mode and utilization threshold

In this section, we explore the best approach to determine the optimal pair of $\{m_{co}, u^{th}_{pm}\}$ minimizing the multi-objective function, J_{dc}. Since external conditions, i.e., outside temperature and user requests, are time-varying, the optimal pair of $\{m_{co}, u^{th}_{pm}\}$ varies as well. Thus, we periodically adjust $\{m_{co}, u^{th}_{pm}\}$ based on the predictions of the external conditions and the predictive sequence of cooling mode transition. Assuming the ideal server consolidation at a certain instant, we can approximate the problem in Section III as follows:

$$\text{Find} \quad \chi(k) = \{m_{co}(k), u^{th}_{pm}(k)\} \tag{11}$$

$$\text{Min} \quad J_{dc}(k) = \sum_{l=k}^{k+N_h-1} \alpha^{l-k}\Big(\widetilde{P}_{cl}(l) + \widetilde{P}_{co}(l) + \widetilde{O}_{tr}(l)\Big) \tag{12}$$

$$\text{s.t} \quad u^{th}_{pm}(l) \geq \frac{\hat{U}_{tot}(l)}{N_{pm}}, \forall l \in [k, k+N_h-1] \tag{13}$$

$$u^{th}_{pm}(l) \leq min\big(u^{max}_{pm}, u^{temp,max}_{pm}(l)\big), \forall l \tag{14}$$

[a]In order to reduce the solution complexity, we find the solution assuming that the ideal server consolidation. The approach is optimistic in that the estimated power consumption is lower than actual scenario due to the fragmentation of the server utilization caused by different utilizations among VMs and fractional ratio of the obtained server utilization to VM utilization in actual scenario. In our future work, we will develop a method which takes into account the effect of the fragmentation to improve the solution quality.

where N_h is the number of time periods; α is a weighting factor, $0 \leq \alpha \leq 1$; $\widetilde{P}_{cl}(l)$, $\widetilde{P}_{co}(l)$, and $\widetilde{O}_{tr}(l)$ are predictions of P_{cl}, P_{co}, and O_{tr} at the l-th period, which are expressed as follows:

$$\widetilde{P}_{cl}(l) = \sum_{mode \in \{act, idle, sleep\}} \widetilde{N}^{mode}_{pm}(l)\widetilde{P}^{mode}_{pm}(l) \tag{15}$$

$$\widetilde{P}_{co}(l) = \big(PUE(u^{th}_{pm}(l)) - 1\big) \cdot \widetilde{P}_{cl}(l) \tag{16}$$

$$\widetilde{O}_{tr}(l) = w^{co}_{tr} \cdot (m_{co}(l) - m_{co}(l-1))^2 \tag{17}$$

where $\widetilde{P}^{mode}_{pm}(l)$ is the estimated average power consumption of server at the l-th period when the operating mode of the server is active (i.e., $u_{pm} = u^{th}_{pm}(k)$ based on the assumption of ideal server consolidation), idle, and sleep modes, and $\widetilde{N}^{mode}_{pm}(l)$ is the corresponding number of servers. PUE is obtained using Eqns. (1) and (5) from Section II. $(m_{co}(l) - m_{co}(l-1))^2$ represents whether cooling mode is switched at the l-th period, and w^{co}_{tr} is a weighting factor which models the overhead caused by cooling mode transition. $\widetilde{N}^{act}_{pm}(l)$, $\widetilde{N}^{idle}_{pm}(l)$, and $\widetilde{N}^{sleep}_{pm}(l)$ are defined as follows:

$$\widetilde{N}^{act}_{pm}(l) = \frac{\widetilde{U}_{tot}(l)}{u^{th}_{pm}(l)} \tag{18}$$

$$\widetilde{N}^{idle}_{pm}(l) = \frac{\hat{U}_{tot}(l)}{u^{th}_{pm}(l)} - \widetilde{N}^{act}_{pm}(l) \tag{19}$$

$$\widetilde{N}^{sleep}_{pm}(l) = N_{pm} - (\widetilde{N}^{act}_{pm}(l) + \widetilde{N}^{idle}_{pm}(l)) \tag{20}$$

where N_{pm} is the number of servers; $\widetilde{U}_{tot}(l)$ is the prediction of average user requests normalized with respect to the maximum number of user requests processed by single server, i.e., $0 \leq \widetilde{U}_{tot}(l) \leq N_{pm}$; $\hat{U}_{tot}(l)$ is the normalized maximum[b] user requests which is characterized a priori based on extensive characterization.

The first constraint (Eqn. (13)) represents the lower bound of $u^{th}_{pm}(l)$ which is determined such that $\hat{U}_{tot}(l)$ user requests can be processed while satisfying SLA requirement. The second constraint (Eqn. (14)) represents the upper bound of $u^{th}_{pm}(l)$, which is determined by the minimum value between the utilization level where multiple VMs can run in single server without acceptable performance loss (i.e., u^{max}_{pm} presented in Section II-A) and the highest utilization satisfying maximum temperature constraint, i.e., $u^{temp,max}_{pm}(l)$ which is calculated based on Eqns. (1) and (3) as follows:

$$u^{temp}_{pm}(l) = m_{co}u^{free,max}_{pm} + (1 - m_{co})u^{elec,max}_{pm} \tag{21}$$

$$u^{mode,max}_{pm}(l) = \frac{T^{max}_{pm} - T^{mode}_{room} - P^{dyn}_{pm}R^{base}_{pm}}{P^{dyn}_{pm}R_{pm}} \tag{22}$$

[b]In this work, we target the SLA violation to be less than 5%. Thus, we used 95th-percentile value instead of the maximum value to characterize the worst-case behavior of the corresponding period. Considering the correlation among VMs, we can use lower percentile values, e.g., 90-, 80-th percentile, etc., to reduce more power consumption while satisfying SLA requirement, as presented in [28].

where *mode* represents cooling mode, i.e., *elec* or *free*; T_{room}^{mode} is server room temperature at corresponding cooling mode.

At the start of k-th period, we solve the optimization problem with two steps: 1) prediction of the external condition, i.e., $\widetilde{U}_{tot}$ and T_{out} for $[k, k+N_h-1]$-th periods and 2) optimization to find $\{m_{co}(k), u_{pm}^{th}(k)\}$.

A. Temperature and workload prediction

At the start of k-th period, we predict $T_{out}(l)$ and $\widetilde{U}_{tot}(l)$ where $k \leq l \leq (k+N_h-1)$. Prediction of T_{out}'s can accurately be predicted by daily and weekly weather forcast. However, accurate prediction of $\widetilde{U}_{tot}$'s is not trivial due to uncertain and non-stationary characteristics of user requests. Fpr accurate prediction, we adopt non-stationary Kalman filter [29], which outperforms other predictors especially when a prediction value is uncertain and non-stationary.

$\widetilde{U}_{tot}(k)$ is predicted based on the history of measured U_{tot} in past few periods as well as the history of the same period in past few days (or weeks). The predictions obtained from the former history is denoted as $\widetilde{U}_{tot}^{(1)}(k)$ while the other is denoted as $\widetilde{U}_{tot}^{(2)}(k)$. Then, we can obtain $\widetilde{U}_{tot}(k)$ by a weighted sum of $\widetilde{U}_{tot}^{(1)}(k)$ and $\widetilde{U}_{tot}^{(2)}(k)$ as follows:

$$\widetilde{U}_{tot}(k) = w_p^{(1)}\widetilde{U}_{tot}^{(1)}(k) + (1-w_p^{(1)})\widetilde{U}_{tot}^{(2)}(k) \qquad (23)$$

where weight, $w_p^{(i)}(k)$ is weight factor

B. Proposed multi-objective optimization

To solve the multi-objective problem considering the uncertainty of T_{out} and $\widetilde{U}_{tot}$, we adopt receding horizon control scheme as follows. At the start of the k-th period, we first predict $\widetilde{U}_{tot}$'s and T_{out}'s for $[k, k+N_h-1]$-th periods as explained in Section IV-A. Second, we find the optimal utilization threshold corresponding to each cooling mode, i.e., $m_{co} = \{0, 1\}$, for $[k, k+N_h-1]$-th periods, as follows. For a given cooling mode, we can express $\widetilde{P}_{dc}(k) = \widetilde{P}_{dc}(k) + \widetilde{P}_{cl}(k)$ as a continuous form with respect to $u_{pm}^{th}(k)$ using Eqns. (15)−(20). In addition, $\widetilde{P}_{dc}(k)$ is convex with respect to $u_{pm}^{th}(k)$ because, as $u_{pm}^{th}(k)$ increases, $\widetilde{P}_{cl}(k)$ is monotonically decreased (due to the decreased number of active servers) while $\widetilde{P}_{co}(k)$ increases because PUE is monotonically increased. Owing to the continuity and convexity of $\widetilde{P}_{dc}(k)$ with respect to $u_{pm}^{th}(k)$ for given $m_{co}(k)$, the unconstrained optimal solution of $u_{pm}^{th}(k)$ can be obtained by finding value which satisfies following equation.

$$\text{Find } u_{pm}^{th}(k) \Longrightarrow \frac{\partial\Big(P_{cl}(k) + P_{co}(k)\Big)}{\partial u_{pm}^{th}(k)} = 0 \qquad (24)$$

The root can be efficiently obtained by root-finding algorithms, e.g., Newton-Raphson method, binary search, etc. [30]. When $u_{pm}^{th}(k)$ satisfies the constraint, we directly set utilization threshold with $u_{pm}^{th}(k)$; otherwise, we set $u_{pm}^{th}(k)$ with lower-bound (Eqn. (13)) and upper-bound (Eqn. (14)) values so as to satisfy the constraint.

Third, with the pairs of $\{m_{co}, u_{pm}^{th}\}$'s and including the overhead caused by cooling mode transition, i.e., O_{tr}, we find the optimal sequence of cooling mode transition from k-th to $(k+N_h-1)$-th periods, i.e., $\chi_{dc}(k) \rightarrow \chi_{dc}(k+1|k) \rightarrow \cdots \rightarrow \chi_{dc}(k+N_h-1|k)$ where $\chi_{dc}(k+l|k)$ is the optimal solution at $(k+l)$-th period when $\chi_{dc}(k)$ is determined as the optimal solution at k-th period. Then, we select only $\chi_{dc}(k)$ and discard the other steps of the sequence. Finally, the entire process is repeated at the start of $(k+1)$-th period with the updated predictions.

V. Experimental Results

A. Experimental setup and compared power optimization methods

We implemented the proposed scheme in CloudSim [31] which is an event-driven simulator providing toolkits to support both system and behavior modeling of Cloud system components such as datacenter, virtual machines (VMs), and resource provisioning policies. We configured the target datacenter with 100homogeneous servers each of which consumes 130W. We created 100 homogeneous VMs all of which process same application. We assumed that VM migration takes 100 sec and 10% performance degradation [15]. We obtained traces of user requests from [26]. We equally distributed the user requests VMs, and we use temperature data measured at EPFL in Lausanne, Switzerland from Jan. 2008 to July 2008.

In our experiments we evaluated the power figures of the following cooling mode decision methods for datacenters:

- **FIXED-TEMP**: a conventional cooling mode decision scheme which uses free cooling only when T_{out} is lower than fixed pre-defined temperature, i.e., $T_{th} = 10°C$ [12], and sets u_{pm}^{th} to u_{pm}^{max}.
- **P-ADAPTIVE**: this is our first proposed scheme which adaptively adjusts the cooling mode and the utilization threshold such that only power consumption of datacenter is minimized.
- **PT-ADAPTIVE**: this is our second proposed scheme which jointly optimizes the power consumption and transition overhead caused by cooling mode transition with receding horizon control scheme.

With the solutions obtained above, we applied the same server consolidation method presented in [19] to the three methods. Note that this paper focuses on the global decision of datacenter, i.e., the pair of cooling mode and corresponding utilization threshold. Thus, we simply compared the three methods above due to the lack of previous works in this optimization topic. The proposed solution is also complementary with exiting power management solutions utilizing dynamic voltage/frequency scaling (DVFS) and VM assignment presented in [14]–[19].

TABLE I
COMPARISONS OF POWER CONSUMPTION AND NUMBER OF COOLING MODE TRANSITIONS IN MAY, JUNE, AND JULY

Period	FIXED-TEMP	P-ADAPTIVE	PT-ADAPTIVE
May 1 $\sim$ May 4	1.00 / 7	0.781 / 8	0.784 / 6
June 1 $\sim$ June 4	1.00 / 0	0.738 / 8	0.743 / 4
July 1 $\sim$ July 4	1.00 / 0	0.879 / 29	0.898 / 13

B. Results

Table I shows comparisons of the three methods in terms of power savings and number of cooling mode transition during the first four days in May, June, and July. Before May, free cooling can be used throughout days because the temperature is mostly lower than the threshold temperature of FIXED-TEMP, i.e., $T_{th} = 10°C$. Thus, we simply provide comparisons during May, June, and July when temperature is mostly higher than $10°C$. The first column is time period we simulated. The second to fourth columns show the normalized power consumption of each method with respect to MAX-UTIL and the number of transitions of cooling mode in each month.

In May, PT-ADAPTIVE provides 21.6% power savings compared to FIXED-TEMP. The reason of the power saving can be analyzed with the traces of cooling mode and utilization schedules presented in Fig. 3. Figs. 3(a) and (b) correspond to the traces for FIXED-TEMP and PT-ADAPTIVE during the first four days in May, respectively. X-axis represents date (month / date). Left and right Y-axis represent cooling mode/utilization and outside temperature, respectively. The outside temperature in May ranges $7\sim22°C$. In FIXED-TEMP (Fig. 3(a)), the utilization threshold is always set to the maximum utilization level, i.e., $u_{th}^{max} = 0.8$, so that as many VMs as possible are consolidated into each active server, and free cooling is applied only when the outside temperature is lower than the threshold temperature, i.e., $10°C$. However, in PT-ADAPTIVE, the time period during which free cooling is used can be much extended by capping maximum power consumption of servers achieved by scheduling the utilization threshold in accordance with the amount of demanding workload as shown in Fig. 3(b).

As shown in Table I, in June, PT-ADAPTIVE provides higher power savings, i.e., 25.7%, compared to May. The reason for the higher power savings is that the outside temperature in June is always higher than $10°C$. Thus, free cooling cannot be used in FIXED-TEMP while free cooling can be still used in PT-ADAPTIVE even when the outside temperature is higher than $10°C$ by adaptively lowering the utilization threshold when the workload of datacenter is lower than the maximum level. On the contrary, the temperature in July is too high, i.e., $14\sim30°C$. Such high temperature makes it infeasible to use free cooling to cool down servers under SLA requirements, which leads to relatively smaller power savings in July, i.e., 10.2%, compared to other months.

In addition, in comparison to P-ADAPTIVE, PT-ADAPTIVE provides almost similar (or slightly less) power savings. However, it provides drastic reduction in the number of cooling mode transitions, thereby, stable cooling mode schedule. In particular, in July, P-ADAPTIVE cannot be used because it switches between cooling modes too often, i.e., seven times per day approximately (=29 times/4 days). Unfortunately, these frequent cooling mode transition causes the adversary effects on power consumption and performance as well as reliability of cooling system. However, by applying the predictive control approach we have proposed in Section IV, i.e., PT-ADAPTIVE, we can reduce the transition down to only 3.25 times per day (=13 times/4 days). Figs. 4 (a) and (b) show the traces for P-ADAPTIVE and PT-ADAPTIVE during the first four days in July, respectively.

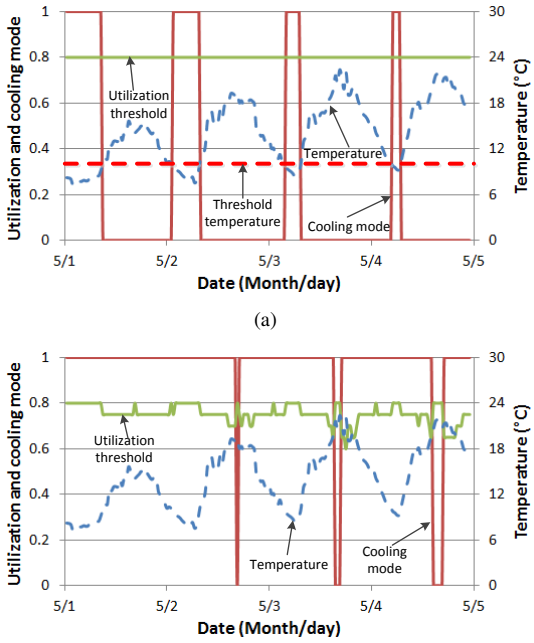

Figure 3. Schedule of cooling mode and utilization threshold in May: (a) MAX-UTIL and (b) PT-ADAPTIVE.

An additional observation is that, the efficiency of the proposed solution truly depends on the energy proportionality of server. Therefore, to explore the effectiveness of the proposed solution, we conducted experiments with various values of power-proportionality of servers, i.e., P_{static}/P_{tot}={0.3,0.5,0.7}. Table II shows the normalized power and the number of transitions in June $1\sim4$. As shown in Table II, our proposed approach obtains more power savings as P_{static}/P_{tot} is lowered. As a matter of fact, when P_{static}/P_{tot} is low, we can use free cooling for longer periods of time by lowering the server utilization threshold, thereby, we have a smaller number of active servers. Furthermore, as state-of-the-art servers are designed to achieve higher energy-proportionality [32], these experiments demonstrate that our proposed approach will be able to provide even more power savings in future datacenter setups.

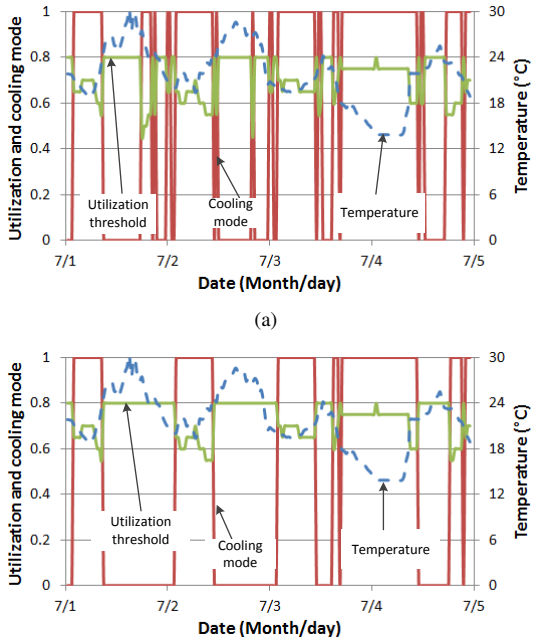

Figure 4. Schedule of cooling mode and utilization threshold in July: (a) P-ADAPTIVE and (b) PT-ADAPTIVE.

TABLE II
COMPARISONS OF POWER CONSUMPTION AND THE NUMBER OF COOLING MODE TRANSITIONS AS P_{static}/P_{tot} CHANGES IN JUNE

P_{static}/P_{tot}	MAX-UTIL	P-ADAPTIVE	PT-ADAPTIVE
0.3	1.00 / 0	0.722 / 2	0.722 / 2
0.5	1.00 / 0	0.738 / 8	0.743 / 4
0.7	1.00 / 0	0.852 / 24	0.878 /12

VI. CONCLUSION

In this paper we have proposed a novel power management approach to reduce the power consumption of datacenters equipped with hybrid cooling architectures by jointly adapting the cooling mode (electrical vs. free cooling) and utilization threshold of servers considering the time-varying nature of the outside temperature and workload characteristics in datacenters. Therefore, we have first analytically formulated the optimization problem. Then, we have proposed an optimization method based on receding horizon control such that the power consumption of datacenter and the overhead caused by switching cooling mode are jointly minimized while satisfying SLA and temperature requirements. According to our experiments, under the climate condition in Lausanne, Switzerland, the proposed scheme can yield up to 25.7% energy savings compared to conventional cooling mode decision scheme which uses free cooling only when the outside temperature is lower than a specific predefined threshold value. In addition, our experimental results have shown that the proposed power minimization approach can provide more power savings as servers expose higher energy-proportionality figures, which outlines its potential with the next-generation servers.

REFERENCES

[1] US Department of Energy, "FEMP Best practices guide for energy-efficient data center design," 2011.
[2] Energy Start Program, US EPA, "Report to Congress on Server and Data Center Energy Efficiency Opportunities," 2007.
[3] J. Choi, *et al.*, "Evaluation of air management system's thermal performance for superior cooling efficiency in high density data centers," in *Elsevier Energy and Buildings*, vol. 43, no. 9, 2011.
[4] E. Pakbaznia, *et al.*, "Temperature-aware dynamic resource provisioning in a power-optimized datacenter," in *Proc. DATE*, 2010.
[5] A. Woods, "Cooling the data center," in *Communications of the ACM*, vol. 53, no. 4, Apr. 2010.
[6] "Google data center," http://www.google.cim/about/datacenters/#.
[7] M. K. Patterson, *et al.*, "Evaluation of air-side economizer use in a compute-intensive data center," in *ACME InterPack*, 2009.
[8] M. Pawlish, *et al.*, "Free cooling: a paradigm shift in data centers," in *Proc. ICIAFs*, 2010.
[9] M. Pervila and J. Kangasharju, "Running servers around zero degrees," in *Proc. GreenNetworking*, 2010
[10] Intel Information Technology, "Reducing data center energy consumption with wet side economizer," in 2007.
[11] Intel Information Technology, "Reducing data center cost with an air economizer," in 2008.
[12] T. Lu, *et al.*, "Investigation of air management and energy performance in a data center in Finland: case study," in *Proc. ENB* 2011.
[13] D. Meisner, *et al.*, "Power management of online data-intensive services," in *Proc. ISCA* 2011.
[14] P. Barham, *et al.*, "Xen and the art of virtualization," in *Proc. SOSP* 2003.
[15] C. Clark, *et al.*, "Live migration of virtual machines," in *Proc. NSDI* 2005.
[16] D. Kusic, *et al.*, "Power and performance management of virtualized computing environments via lookahead control," in *Cluster Comput, Springer* 2009.
[17] G. Dhiman, *et al.*, "vGreen: a system for energy efficient computing in virtualized environments," in *Proc. ISLPED* 2009.
[18] J. Xu *et al.*, "Multi-objective virtual machine placement in virtualized data center environments," in *Proc. CPScom*, 2010.
[19] J.-W. Jang, *et al.*, "Energy reduction in consolidated servers through memory-aware virtual machine scheduling," , in *IEEE Transactions on Computures*, vol. 60, no. 4, Apr. 2011.
[20] P. Bodik, *et al.*., "A case for adaptive datacenters to conserve energy and improve reliability." in *Technical Report No. UCB/EECS-2008-127* 2008.
[21] S. Rivoire, *et al.*. "A Comparison of High-Level Full-System Power Models," in *Proc. HotPower* 2008.
[22] D. Economou, *et al.*, "Full-system power analysis and modeling for server environments," in *Proc. MoBS* 2006.
[23] W. Huang, *et al.*, "HotSpot: a compact thermal modeling methodology for early-stage VLSI design," in *IEEE TVLSI* vol. 14, no. 5, pp. 501-513, 2006.
[24] D. C. Hwang, *et al.*, "Energy savings achievable through liquid cooling," in *Proc. ITherm* 2010.
[25] T. J. Breen, *et al.*, "From chip to cooling tower data center modeling: Part I Influence of server inlet temperature and temperature rise across cabinet," in *Proc. ITherm* 2010.
[26] T. Benson, *et al.*, "Understanding data center traffic characteristics," in *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 1, Jan. 2010.
[27] R. Carroll, *et al.*, "Dynamic optimization solution for green service migration in data centres," in *Proc. IEEE ICC* 2011.
[28] A. Verma, *et al.*, "Server workload analysis for power minimization using consolidation," in *Proc. USENIX* 2009.
[29] S.-Y. Bang, *et al.*, "Run-time adaptive workload estimation for dynamic voltage scaling," in *IEEE TCAD*, vol. 28, no. 9, pp. 1334-1347, Sep. 2009.
[30] K. Madsen, "A root-finding algorithm based on Newton's method," in *Mathmatics and statistics, Springer*, vol. 13, no. 1, 1973.
[31] R. Buyya, *et al.*, "Modling and simulation of scalable cloud computing environment and the CloudSim toolkit: challenges and opportunities," in *Proc. HPCS* 2009.
[32] D. Meisner, *et al.*, "PowerNap: eliminating server idle power," in *Proc. ASLPOS* 2009.